Stochastic determination of matrix determinants

Sebastian Dorn[*] and Torsten A. Enßlin

Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany
and Ludwigs-Maximilians-Universität München, Geschwister-Scholl-Platz 1, D-80539 München, Germany

Matrix determinants play an important role in data analysis, in particular when Gaussian processes are involved. Due to currently exploding data volumes, linear operations—matrices—acting on the data are often not accessible directly but are only represented indirectly in form of a computer routine. Such a routine implements the transformation a data vector undergoes under matrix multiplication. While efficient probing routines to estimate a matrix's diagonal or trace, based solely on such computationally affordable matrix-vector multiplications, are well known and frequently used in signal inference, there is no stochastic estimate for its determinant. We introduce a probing method for the logarithm of a determinant of a linear operator. Our method rests upon a reformulation of the log-determinant by an integral representation and the transformation of the involved terms into stochastic expressions. This stochastic determinant determination enables large-size applications in Bayesian inference, in particular evidence calculations, model comparison, and posterior determination.

I. MOTIVATION

Current and future physical observations generate huge data streams to be analyzed. Particle physics, biophysics, astronomy, and cosmology are representatives of current scientific fields of interest that are undergoing a revolution driven by increasing data volume. Typical large data sets in cosmology are, for instance, observations of the cosmic microwave background [1,2] as well as of the large-scale structure [3–5] as they are often wide- or all-sky observations carried out by telescopes with remarkable resolution. In order to extract information about the universe or physics in general, Bayesian inference methods becomes more and more frequently used as their large computational demands become more feasible thanks to technology developments. The signal of interest to be extracted from data could be almost everything, ranging from just a single parameter (e.g., the level of local non-Gaussianity of the cosmic microwave background [6,7]) to a full four-dimensional reconstruction of the structure growth in the universe [8,9]. Such ambitious Bayesian analyses often invoke linear transformations of the data or of estimated signal vectors.

The size of the involved data and signal spaces often bans the explicit representation of matrices acting on these spaces by their individual matrix elements. A prominent example appearing in many analyses is, for instance, the covariance matrix of a multivariate Gaussian distribution of a vector valued quantity, which describes the two-point correlation structure of the said quantity. Due to their large dimensions such matrices are often only representable by a computer routine, which implements the application of the matrix to a vector without storing or even calculating the individual matrix elements. Such routines often invoke fast Fourier transformations and other efficient operations, which in combination render nonsparse matrices into easily computable basis systems. We refer to such a matrix as an *implicit matrix*. For instance, calculating the model evidence often requires calculating determinants of such matrices. This work provides an efficient way to numerically calculate determinants given only by an implicit matrix representation.

The remainder of this work is organized as follows: In Sec. II we introduce the formalism of the stochastic estimation of an implicit matrix and present two numerical examples. Section III provides a perspective of possible applications in science. Results are summarized in Sec. IV.

II. PROBING THE LOG-DETERMINANT OF AN IMPLICIT MATRIX

A. Formalism

Let $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ be an implicitly defined, complex-valued, square matrix of order n. Implicitly means that the particular entries of the matrix are not accessible, for instance, if dealing with large data sets, where an explicit storage of A might exceed the memory of the computer. However, the action of the matrix as a linear operator is assumed to be known and given by a computer routine implementing the mapping $x \mapsto Ax$.

Motivated by applications in science and statistics (Secs. I and III), in particular by signal reconstruction techniques and model comparison in astronomy and cosmology, where the determinant of a covariance matrix is required (Sec. III), we constrain the variety of different types of matrices by requesting that the matrix A of interest is either *weak diagonal dominant* or *Hermitian positive definite*. The term *weak diagonal dominant* is defined by

$$|a_{ii}| \geqslant \sum_{i \neq j} |a_{ij}| \quad \forall i, \tag{1}$$

while *Hermitian positive definite* means

$$A^{\dagger} = A \quad \text{and} \quad x^{\dagger} A x > 0 \quad \forall x \in \mathbb{C}^n \backslash \{0\} \tag{2}$$

with † denoting the adjoint.

The diagonal and the trace of an implicit matrix can be obtained by exploiting common probing routines [10–13]. A stochastic estimate of the diagonal of the linear operator A is

───────
[*]sdorn@mpa-garching.mpg.de

given by

$$\mathrm{diag}(A) = \langle \xi \star A\xi \rangle_{\{\xi\}} \approx \frac{1}{M}\sum_{i=1}^{M} \xi_i \star A\xi_i, \qquad (3)$$

where $\star$ denotes a componentwise product, $M = |\{\xi\}|$ the sample size, and $\langle \cdot \rangle_{\{\xi\}}$ the arithmetic mean over ξ with $M \to \infty$. The probing vectors $\xi \in \mathbb{C}^n$ are random variables, whose components $x(x')$ fulfill the condition $\langle \xi_x \xi_{x'} \rangle_{\{\xi\}} = \delta_{xx'}$. Analogously to the diagonal of an operator its trace can be probed by, e.g.,

$$\mathrm{tr}(A) = \langle \xi^\dagger A\xi \rangle_{\{\xi\}}. \qquad (4)$$

Recently, there have been investigations to improve these straightforward probing methods by exploiting Bayesian inference [10]. This has been achieved by reformulating the process of stochastic probing of an operator's diagonal (trace) as a signal inference problem. As a result, it requires fewer probes than the purely stochastic methods and thus can decrease the computational costs. With the phrase *operator probing*, be it trace or diagonal probing, we subsequently refer to the entirety of probing methods in general.

The linear operator A can be split into a diagonal matrix $D \in \mathbb{C}^{n\times n}$ and a matrix $N \in \mathbb{C}^{n\times n}$, which contains the off-diagonal part of A, i.e.,

$$A = D + N. \qquad (5)$$

We are now interested in the value of its determinant or of its log-determinant, $\Delta \equiv \ln[\det(A)]$. In case A is mainly dominated by its diagonal (i.e., $ND^{-1} \ll \mathbb{1}$ spectrally), a Taylor expansion of the log-determinant might be a reasonable approximation,

$$\begin{aligned}\Delta &= \ln[\det(D+N)]\\ &= \ln[\det(D)] + \mathrm{tr}[ND^{-1}] + \mathcal{O}(\mathrm{tr}[(ND^{-1})^2]), \quad (6)\end{aligned}$$

which is sometimes feasible dealing with implicit operators, e.g., see Refs. [7,14] for recent applications in cosmic microwave background physics. This approximation, however, breaks down when the relation $ND^{-1} \ll \mathbb{1}$ (spectrally) is violated. In order to circumvent this problem we introduce the quantity

$$\Delta(t) \equiv \ln[\det(D+tN)] \qquad (7)$$

with the pseudotime parameter $t \in [0,1]$. For a sufficiently small t the approximation of Eq. (6) becomes valid. This property can be used together with a few mathematical manipulations (for details see Appendix A) to obtain the formula

$$\begin{aligned}\Delta &= \int_0^1 dt\, \mathrm{tr}[N(D+tN)^{-1}] + \Delta(0)\\ &= \int_0^1 dt\, \langle \xi^\dagger N(D+tN)^{-1}\xi \rangle_{\{\xi\}} + \Delta(0) \qquad (8)\end{aligned}$$

that represents a stochastic estimate of the log-determinant of A using operator probing. In particular, the following steps are required to evaluate Eq. (8):

(1) Diagonal (operator-) probing to split A into

$$A = \underbrace{\mathrm{diag}(A)}_{\equiv D} + \underbrace{A - \mathrm{diag}(A)}_{\equiv N},$$

(2) an approach to invert $D + tN$ in Eq. (8), e.g., the *conjugate gradient* method [15],

(3) trace (operator-) probing to evaluate the integrand, and

(4) a numerical integration method, e.g., applying Simpson's rule.

It might immediately strike the eye of the reader that one recaptures the simple first-order Taylor-expanded version of the log-determinant, Eq. (6), when dropping the pseudotime dependency of the integrand in Eq. (8) by requesting $t = 0$. This means that in case of dealing with diagonal dominant operators the value of the correct log-determinant might be received by a coarse numerical integration since the integrand close to $t = 0$ already yields the main correction, which might decrease the computational costs, see Sec. II B.

Equation (8) further represents the main result of this paper and can be regarded as a special case of calculating partition functions (see Sec. III and Refs. [16,17]). Although the first line of it, the integral representation of the log-determinant, was also, independently of our work, found by mathematicians 10 years ago [18], it is (to our knowledge) not known in the community of physics or signal inference. The connection to stochastic estimators, however, is a novel way to evaluate the log-determinant of implicitly defined matrices that enables previously impossible calculations, see Sec. III.

B. Numerical example

We address here a simple and also exactly solvable numerical example referring to (Bayesian) signal inference problems or, in general, statistical problems in physics (see Secs. III A and III B), where the log-determinant of a covariance matrix A is of interest. If we assume statistical isotropy and homogeneity of a physical field, its covariance matrix can be parametrized by a so-called power spectrum. This is often a reasonable assumption,[1] e.g., in astronomy and physical cosmology, when applying the cosmological principle. In this case, the covariance matrix becomes diagonal in Fourier space,

$$A_{kk'} = c_k \delta_{kk'}, \qquad (9)$$

with respective Fourier modes k, k' and power spectrum c_k. It is straightforward to show that the position space representation of $A_{kk'}$, given by $A_{xx'} = \mathcal{F}^\dagger_{xk} A_{kk'} \mathcal{F}_{k'x'}$ with Fourier transformation $\mathcal{F}$, is nondiagonal if and only if $c_k \neq \mathrm{const} \;\; \forall k$. In order to apply the stochastic estimator of the log-determinant we use two special forms of the power spectrum, given by

$$c_k = \frac{1}{(1+k)^\alpha} \qquad (10)$$

with α set to 2 or 4. A value of $\alpha = 2$ describes a mostly diagonal dominant matrix, whereas $\alpha = 4$ exhibits a significant

---

[1]Referring to Bayesian evidence calculations such a matrix might be the prior or posterior covariance, see Sec. III for details.
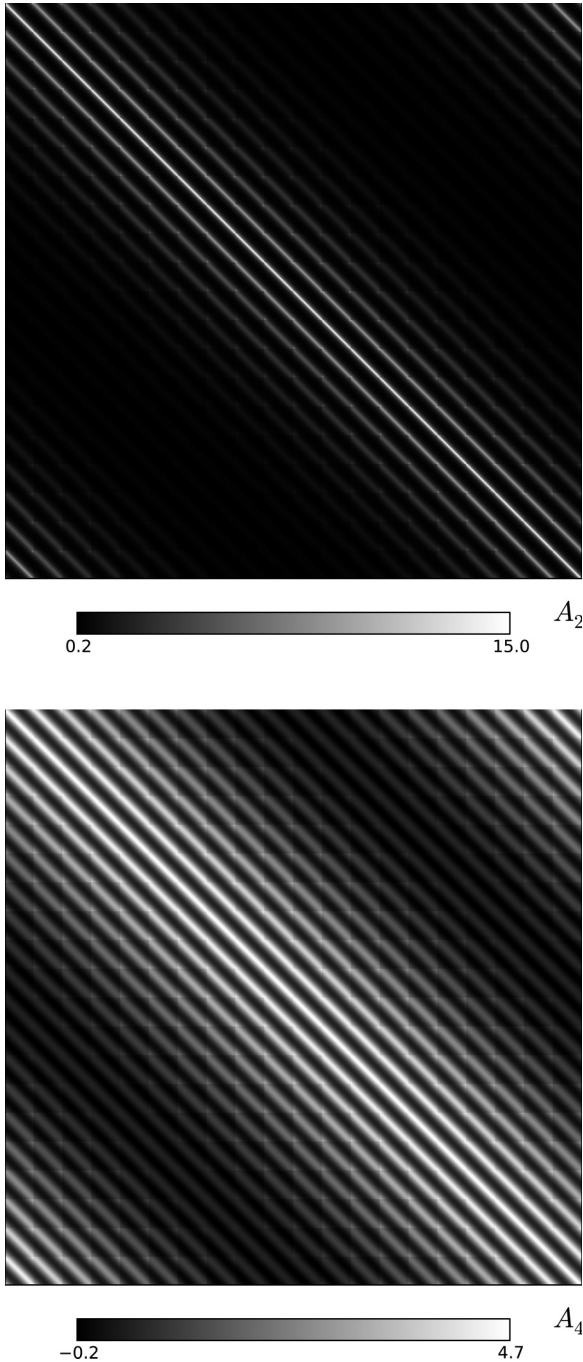
FIG. 1. Illustration of the matrices A_2 (top) and A_4 (bottom) in position space with linear color bars.

nondiagonal structure in position space. To be precise, in the following we use a regular, two-dimensional, real-valued grid (over $\mathcal{T}^2$) of $n = 20 \times 20$ pixels to represent our position space, resulting in a matrix A consisting of $n \times n = 1.6 \times 10^5$ real numbers. See Fig. 1 for an illustration thereof.

For both matrices, which we refer to as A_2 and A_4, we apply Eq. (8) given an explicit and implicit numerical implementation. For the explicit variant there also exist

well-understood, precise numerical methods[2] to calculate the determinant. Therefore, the numerical results of such a method can be regarded as our gold standard and hence serve as a reference for the probing results. Henceforth we will refer to it using the subscript "correct." Both variants, the explicit and implicit implementation, are realized using the tools of NIFTY [20].

After the separation of A_2 and A_4 into diagonal and off-diagonal parts by applying diagonal probing, we calculate the integrands of Eq. (8) for the m-part-discretized interval of $t \in [0,1]$ by using the conjugate gradient method as well as trace probing and perform the numerical integration afterwards by using Simpson's rule. The operator probing as well as the conjugate gradient method have also been realized using NIFTY. Furthermore, we introduce the quantities

$$\Delta(x) \equiv \int_0^x dt\, \mathrm{tr}[N(D + tN)^{-1}] + \Delta(0), \quad x \in [0,1] \quad (11)$$

to study the convergence to the final value and $\Delta(m)$ to investigate the dependency on the discretization of the integration interval, see Figs. 2, 3, and 4.

We used a rather low sample size of $M = 8$ for trace probing [see Eqs. (4) and (3)] to demonstrate the applicability of the method to large data sets. The discretization of the pseudotime interval into m parts was chosen to be $m = 10^3$ for A_4 and only $m = 10$ for A_2, see in particular Fig. 4, which illustrates the dependence of the probing result on m.

C. Discussion

The exact numerical values of the determinant calculation using explicit and implicit representations of A_4 and A_2 can be found in Table I. The results of the probing method (implicit) compared to the correct and the explicit method, where Eq. (8) can be evaluated without using a conjugate gradient or probing techniques, are accurate for both matrices. It is remarkable that despite using a relatively small sample size of $M = 8$ for the trace probing the absolute errors remain relatively small. The reason for this is that the pseudotime integration over all probed integrands averages the probing error. This is of particular importance when applying the log-determinant probing to large data sets, where a large sampling size should be avoided to save computational time. These errors can be decreased further, of course, by an increase of the sampling size and a refinement of the numerical integration.

The results of the trace (integrand) probing and the determinant's convergence behavior as well as their respective errors with respect to the explicit representation can be found in Figs. 2 and 3. Note that the scaling of the ordinate is logarithmic. For both matrices, but especially for A_4, the largest contribution to the integral of Eq. (8) comes from late t values. Therefore, if dealing with big data sets, one could divide the integration interval not into m equal parts but by starting with a rather coarse discretization for small t values and subsequently refining it for larger values, e.g., by substituting dt by $d\ln(t')$ and thereby saving computational

---

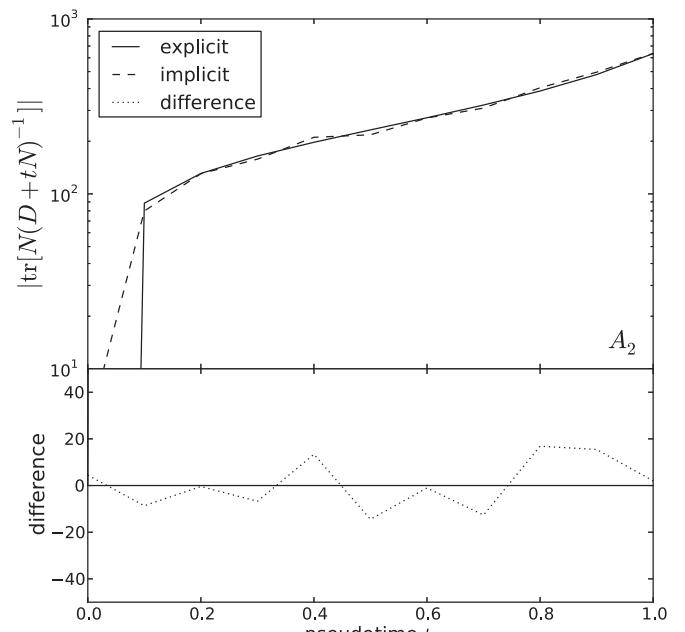[2]See, for instance, the method described at [19], which is based on LU factorization.
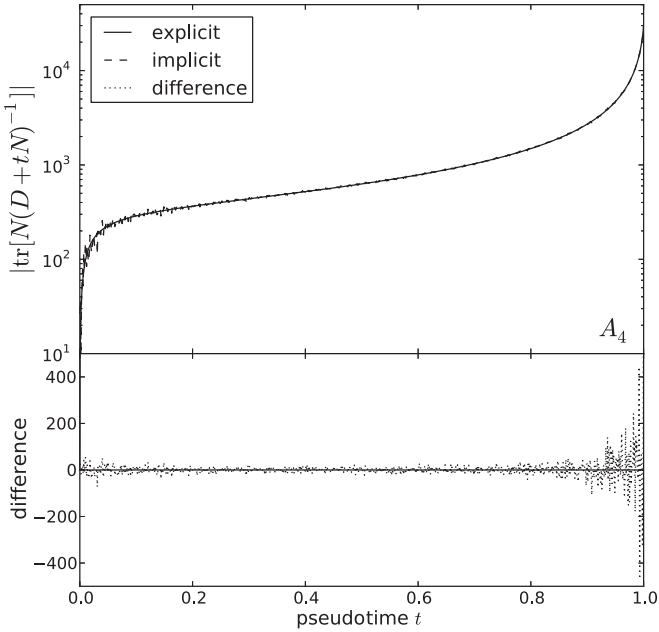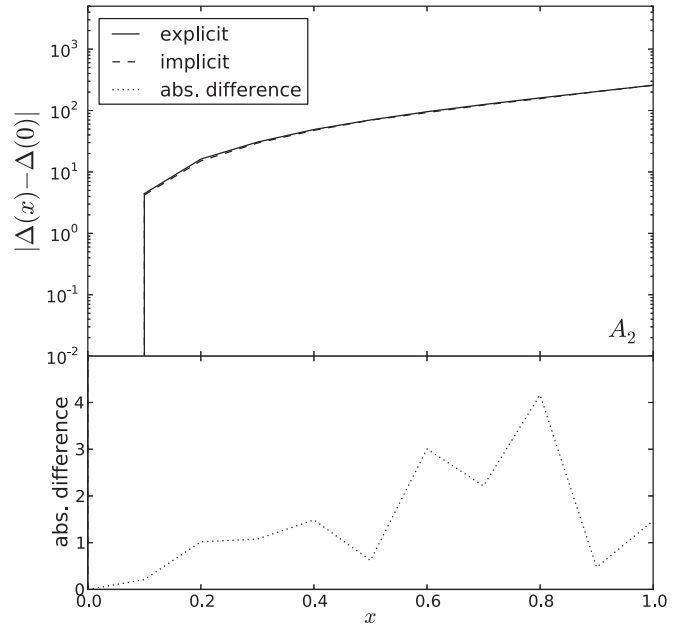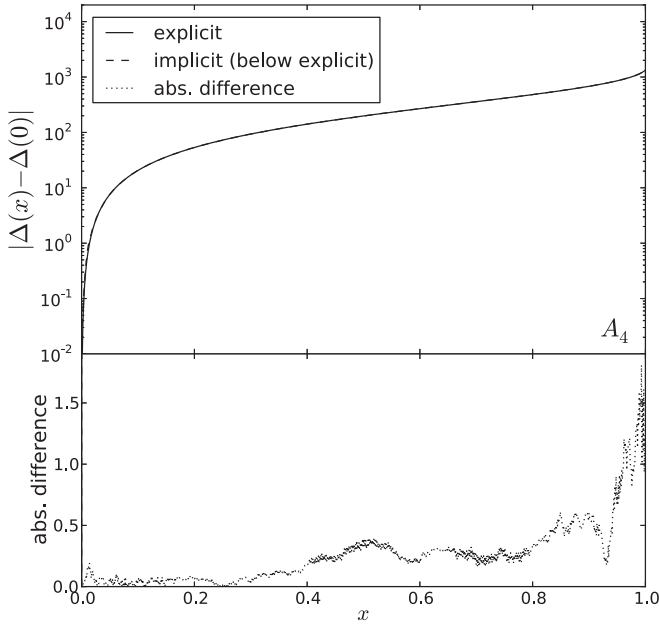
FIG. 2. The integrand of Eq. (8) (lower panel) and $\Delta(x)$ (upper panel) for explicit and implicit representations of A_4.

FIG. 3. The integrand of Eq. (8) (lower panel) and $\Delta(x)$ (upper panel) for explicit and implicit representations of A_2 with only $m = 10$ steps in pseudotime.

costs. This, however, might depend on the particular shape of the matrix and has to be studied case by case.

The dependency of the numerical value of the determinant of A_4 on the discretization (in m equal parts) of the integration interval can be found in Fig. 4 and shows that even a small value of m adds significant corrections to the result. The result for $m \propto \mathcal{O}(10)$ is, for instance, better than just using the determinant of the diagonal, $\Delta(0)$. This might be used in practice to investigate cheaply whether the nondiagonal structure of a matrix influences the determinant significantly.

A huge advantage of the probing method discussed here is the possibility to parallelize the numerical calculation almost completely. To be precise, the diagonal probing beforehand, the pseudotime integral, as well as every single trace probing can be parallelized fully. The only operation that cannot be parallelized is the conjugate gradient method as it is a potential minimizer, using at least the previous step to calculate the next one.

The determination of a suitable choice of the involved parameters m and M as well as the precision parameters for the used conjugate gradient approach and numerical integration method depend highly on the matrix to be studied. The computational costs and precision of the introduced determinant calculation thus depend on the combination of the chosen methods for diagonal and trace probing, numerical
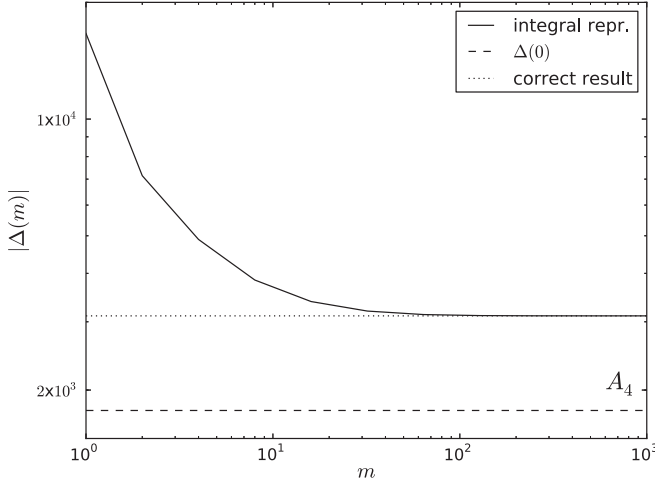
FIG. 4. Dependency of the determinant's result on the discretization of the integration interval into m parts, using A_4.

integration, the method to numerically invert the matrix $D + tN$, and the matrix A itself. Since it is therefore not possible to make general statements we consciously avoid here such a discussion of computational costs and precision with respect to m and M. A more pragmatic way to estimate these parameters would be to downscale the problem of interest until the matrix of interest fits into the memory of the computer and to subsequently perform mock tests to obtain a suitable choice for the parameters discussed above. Afterwards these values can be extrapolated to the size of the real problem.

III. APPLICATIONS IN SCIENCE

Within this section we present a selection of possible applications in science. Although there are a vast number of research fields and topics which might benefit from the stochastic estimation of a log-determinant we focus henceforth on a selection of usages in Bayesian signal inference, in particular in physics and only present simple examples. Exact, more complicated examples can be found in the cited works within this section.

TABLE I. Results of the numerical determinant calculations with and without probing. The absolute errors of the probing method are defined by $\epsilon_1 = |\Delta_{\text{explicit}}(1) - \Delta_{\text{implicit}}(1)|$ and $\epsilon_2 = |\Delta_{\text{correct}} - \Delta_{\text{implicit}}(1)|$. Differences between ϵ_1 and ϵ_2 arise from the discretized, numerical integration.

	A_2	A_4
$\Delta(0)$	-1308.05	-1771.57
Δ_{correct}	-1566.99	-3107.28
$\Delta_{\text{explicit}}(1)$	-1566.81	-3107.29
$\Delta_{\text{implicit}}(1)$	-1565.33	-3108.41
m	10	1000
M	8	8
ϵ_1	1.48	1.12
ϵ_2	1.66	1.13

A. Evidence calculations and model selection

The Bayesian evidence $\mathcal{P}(d)$ is a measure for the quality of the model and hence for all assumed model parameters for the data d [21]. To keep it short and simple we assume a model that describes a linear measurement of a Gaussian signal s with additive, signal-independent, Gaussian noise n, i.e.,

$$d = Rs + n, \tag{12}$$

where R represents a linear operator. A Gaussian distribution of a variable x is defined by

$$\mathcal{P}(x) = \mathcal{G}(x, X) \equiv \frac{1}{\sqrt{|2\pi X|}} \exp\left\{ -\frac{1}{2} x^\dagger X^{-1} x \right\} \tag{13}$$

with related covariance matrix X and mean

$$\langle x \rangle_{\mathcal{P}(x)} \equiv \int \mathcal{D}x \, x \, \mathcal{P}(x). \tag{14}$$

$\int \mathcal{D}[\cdot]$ denotes a phase space integral and $|\cdot|$ the determinant. Under these circumstances the evidence can be calculated as

$$\begin{aligned}
\mathcal{P}(d) &= \int \mathcal{D}s \int \mathcal{D}n \, \mathcal{P}(d, s, n) \\
&= \int \mathcal{D}s \int \mathcal{D}n \, \delta(d - Rs - n) \mathcal{P}(n|s) \mathcal{P}(s) \\
&= \sqrt{\frac{|2\pi C_{s|d}|}{|2\pi C_s||2\pi C_n|}} \exp\left\{ -\frac{1}{2} \left(d^\dagger C_n^{-1} d - j^\dagger C_{s|d} j \right) \right\},
\end{aligned} \tag{15}$$

with

$$j = R^\dagger C_n^{-1} d, \quad C_{s|d}^{-1} = R^\dagger C_n^{-1} R + C_s^{-1}, \tag{16}$$

and the signal and noise covariances C_s and C_n, respectively. Therefore, to calculate the Bayesian model evidence, one often[3] has to calculate determinants of covariance matrices. This might be done by probing [Eq. (8)] if dealing with implicit matrices [last line of Eq. (15)] instead of performing the multidimensional integral [second last line in Eq. (15)] numerically. The latter has been done, for instance, in the field of inflationary cosmology [22,23] by the method of nested sampling [24,25].

This is especially of importance in the field of model selection or comparison [21], where from an observation—the data—one wants to infer which theory reproduces the observation best. Switching from one model to another means, for instance,[4] to exchange R in Eq. (15), which directly affects the determinant containing $C_{s|d}$. Thus, the calculation of the determinant is mandatory here.

---

[3] By the word "often" we refer to cases, in which at least one marginalization [see Eq. (17)] can be performed analytically (approximated with high precision) to obtain a model-dependent determinant.

[4] We focus here on R for simplicity only. One could also, additionally, exchange the prior covariances C_n and C_s, the assumed prior statistics, the parametrization of the data, and so on.

B. Posterior distribution including marginalizations

In the field of signal inference one is typically interested in reconstructing a set of i parameters p^i with uncertainty from some observation, the data d. This information is delivered by the posterior, given by[5] [26] $\mathcal{P}(p^i|d) \propto \mathcal{P}(d|p^i)\mathcal{P}(p^i)$. Often, however, this inference problem is degenerate, caused by a so-called nuisance parameter. For example, consider the calibration of an instrument is of interest and not the signal. In this case the signal s represents the nuisance parameter. The common procedure to circumvent this problem is to marginalize over these parameters,

$$\mathcal{P}(p^i|d) \propto \int \mathcal{D}s \int \mathcal{D}n \mathcal{P}(d,s,n|p^i)\mathcal{P}(p^i). \quad (17)$$

To continue with the simple example of Sec. III A we assume again Gaussian distributions for s and n and a linear measurement but with explicit dependency on p^i, i.e., $d = (Rs)[p^i] + n$. If we further follow the example of calibration, the parameter p^i might be a calibration coefficient, thus affecting only R. This yields $(Rs)[p^i] = R[p^i]s$ and therefore

$$\mathcal{P}(p^i|d) \propto \left\{ \int \mathcal{D}s \mathcal{G}(d - R[p^i]s, C_n)\mathcal{G}(s, C_s) \right\} \mathcal{P}(p^i). \quad (18)$$

This integration can be performed analytically, producing an in general non-Gaussian probability distribution with p^i-dependent normalization (and exponent) similar to Eq. (15),

$$\mathcal{P}(p^i|d) \propto \sqrt{\left|2\pi C_{s|d}[p^i]\right|} \mathcal{P}(p^i) \exp\left\{\tfrac{1}{2} j^\dagger[p^i]C_{s|d}[p^i]j[p^i]\right\}, \quad (19)$$

with $C_{s|d}[p^i]$ and $j[p^i]$ now containing $R[p^i]$ instead of R. In case the covariance matrices or $R[p^i]$ are only given by a computer routine (implicit representation of a matrix) one could use Eq. (8) to probe the determinant.

A variety of scientific fields are affected by this problem. For example, the extraction of the level of non-Gaussianity of the cosmic microwave background [7,14] in cosmology, the problem of self-calibration [27–29] in general, or lensing in astronomy [30].

C. Realistic astronomical example

In order to study a more realistic example we consider a measurement device with spatially constant but unknown calibration amplitude, parametrized by $1 + \gamma \in \mathbb{R}$, scanning a specific patch of the sky. The measured and assumed to be Gaussian sky signal s is affected by the instrument via a convolution $\mathcal{C}$ with a Gaussian kernel of standard deviation $\sigma = 0.05$. Additionally, the observation might be disturbed by fore- and backgrounds. For this reason we include an observational mask M_o, which cuts out 20% of the sky. The noise n is still assumed to be Gaussian and uncorrelated with the signal. Hence, the measurement equation is given by

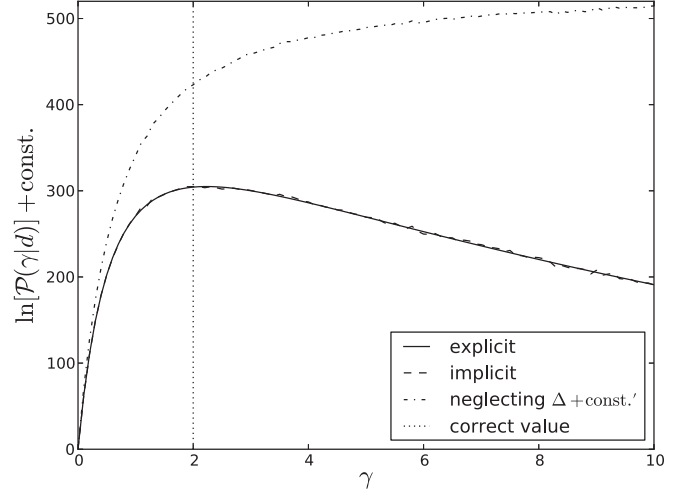$$d = R[\gamma]s + n = (1 + \gamma)M_o\mathcal{C}s + n. \quad (20)$$

FIG. 5. Logarithmic posterior of the calibration amplitude parameter γ using implicit and explicit representations of the involved operators, see Eq. (21) and Eq. (22) for details. The abbreviation Δ denotes the logarithm of the term given by Eq. (22).

To calibrate the measurement device the calibration posterior $\mathcal{P}(\gamma|d)$ has to be determined. The resulting calibration mean $\langle\gamma\rangle_{\mathcal{P}(\gamma|d)}$ can be regarded as an external calibration if the *a priori* knowledge on the signal is sufficiently strong. Otherwise one could infer the signal and calibration amplitude γ simultaneously from data using iterative approaches [28]. Using Eq. (19) as well as a flat prior on γ we obtain

$$\ln\mathcal{P}(\gamma|d) = -\tfrac{1}{2}\ln\left|C_{s|d}^{-1}[\gamma]\right| + \tfrac{1}{2}j^\dagger[\gamma]C_{s|d}[\gamma]j[\gamma] + \text{const}, \quad (21)$$

which exhibits in particular the γ-dependent determinant

$$\left|C_{s|d}^{-1}[\gamma]\right| = \left|(1 + \gamma)\mathcal{C}^\dagger M_o^\dagger C_n^{-1} M_o\mathcal{C}(1 + \gamma) + C_s^{-1}\right|. \quad (22)$$

For the numerical evaluation of Eq. (21) we use the settings of Sec. II B with $C_s(k,k') = (1 + k)^{-3}\delta_{kk'}$, a calibration amplitude parameter of $\gamma = 2$, and a noise covariance of $(C_n)_{x,x'} = 10^{-1}\delta_{xx'}$ to generate a data realization. The pseudotime interval has been discretized into 10^2 parts. The numerically determined calibration posterior for a given data realization can be found in Fig. 5, which demonstrates again the efficiency of the stochastic method using only eight probes for a single trace probing operation. The figure also illustrates the impact of the determinant on the log-posterior, which would not peak in the shown interval without it.

IV. SUMMARY

Motivated by the problem of finding a way to efficiently determine the determinant of an implicitly defined matrix or operator, we derived a formula, Eq. (8), representing a stochastic estimate of its log-determinant. This has been achieved by reformulating the log-determinant as an integral representation and transforming the involved terms into stochastic expressions, which includes a numerical integration and a trace probing. Numerical examples have shown that the discretization of the integration interval may be very coarse in case the probed operator is sufficiently diagonal.

---

[5]Note that in this case the evidence is just a scalar which normalizes the posterior, therefore we merely state proportionalities.

In case it exhibits a significant nondiagonal structure one has to fine-grain the discretization of this interval. The number of probes necessary for the trace probing, however, remains very low in the studied examples. These facts combined with the almost complete parallelizability of this approach might keep the computational costs within reasonable limits in many situations.

This method clearly has more general applications but might in particular be useful for Bayesian signal inference and model comparison when dealing with large data sets as often given, for instance, in astronomy and cosmology. To be precise, it might be beneficial in all fields where the numerical calculation of a determinant of an operator is mandatory.

APPENDIX: INTEGRAL REPRESENTATION OF THE LOG-DETERMINANT OF A MATRIX

Here Eq. (8) is derived. Following Sec. II the log-determinant Δ of an operator A can be parametrized by $\Delta = \ln[\det(D + N)]$ with D being the diagonal and N the off-diagonal part of A. Since Δ can be Taylor-expanded for small N (spectrally compared to D) only, we employ a method from the field of renormalization theory [29,32]. Accordingly, we introduce an expansion parameter $\delta t \ll 1$ to suppress the influence of N. In particular, we replace Δ by $\ln[\det(D + \delta t N)]$ for a moment. For sufficiently small values

of δt, in the following interpreted as tiny pseudotime steps, we can approximate Δ by Eq. (6). Theoretically, a single pseudotime step could be infinitesimal small, enabling the formal definition of the derivative

$$
\begin{aligned}
\frac{d\Delta(t)}{dt} &\equiv \lim_{\delta t \to 0} \frac{\ln[\det(D + (t + \delta t)N)] - \ln[\det(D + t N)]}{\delta t} \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \ln\left[\det(\mathbb{1} + \delta t\, N[D + t N]^{-1})\right] \\
&= \lim_{\delta t \to 0} \frac{1}{\delta t} \mathrm{tr}[\ln(\mathbb{1} + \delta t\, N[D + t N]^{-1})] \\
&= \mathrm{tr}[N[D + t N]^{-1}],
\end{aligned}
\tag{A1}
$$

with the definition

$$
\Delta(t) \equiv \ln[\det(D + t N)].
\tag{A2}
$$

Integrating the pseudotime derivative of $\Delta(t)$ yields the integral representation of the log-determinant,

$$
\Delta = \int_0^1 dt\, \mathrm{tr}[N(D + t N)^{-1}] + \Delta(0).
\tag{A3}
$$

This integral representation has also been found by Ref. [18], where its validity has been proven for weak diagonal dominant and Hermitian positive definite matrices. In particular one has to ensure the existence of the inverse matrix of the integrand of Eq. (A3).

Finally, we replace the trace by stochastic trace probing and perform the pseudotime integral by an numeric integration method. This yields

$$
\Delta = \int_0^1 dt\, \langle \xi^\dagger N(D + t N)^{-1} \xi \rangle_{\{\xi\}} + \Delta(0).
\tag{A4}
$$

[1] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, Astrophys. J **208**, 20 (2013).

[2] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday *et al.*, Planck Collaboration, Astron. Astrophys. **571**, A16 (2014).

[3] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page *et al.*, Astrophys. J. Suppl. Ser. **170**, 377 (2007).

[4] G. J. Hill, K. Gebhardt, E. Komatsu, N. Drory, P. J. MacQueen, J. Adams, G. A. Blanc, R. Koehler, M. Rafal, M. M. Roth, A. Kelz, C. Gronwall, R. Ciardullo, and D. P. Schneider, in *Panoramic Views of Galaxy Formation and Evolution*, edited by T. Kodama, T. Yamada, and K. Aoki, Vol. 399 (Astronomical Society of the Pacific Conference Series, San Francisco, CA, 2008), pp. 115–118.

[5] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly,

J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto,

D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, and N. Yasuda (SDSS Collaboration), Astronomical J. **120**, 1579 (2000).

[6] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005).

[7] S. Dorn, N. Oppermann, R. Khatri, M. Selig, and T. A. Enßlin, Phys. Rev. D **88**, 103516 (2013).

[8] F. S. Kitaura and T. A. Enßlin, Mon. Not. R. Astron. Soc. **389**, 497 (2008).

[9] J. Jasche and B. D. Wandelt, Mon. Not. R. Astron. Soc. **432**, 894 (2013).

[10] M. Selig, N. Oppermann, and T. A. Enßlin, Phys. Rev. E **85**, 021134 (2012).

[11] E. Aune, D. P. Simpson, and J. Eidsvik, Statistics and Computing **24**, 247 (2014).

[12] M. Hutchinson, Commun. Stat. Simul. Comput. **18**, 1059 (1989).

[13] C. Bekas, E. Kokiopoulou, and Y. Saad, Appl. Numer. Math. **57**, 1214 (2007).

[14] S. Dorn, E. Ramirez, K. E. Kunze, S. Hofmann, and T. A. Enßlin, J. Cosmol. Astropart. Phys. 06 (2014) 048.

[15] M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).

[16] M.-D. Wu and W. Fitzgerald, *Bayesian Multimodal Evidence Computation by Adapti Tempering MCMC*, Vol. 79 of Fundamental Theories of Physics (Springer, Amsterdam, 1996).

[17] D. C. M. Dickson, M. R. Hardy, and H. R. Waters, *Physics and Probability—Essays in Honor of Edwin T. Jaynes*, 1st ed. (Cambridge University Press, Cambridge, 2004).

[18] J. Du and J. Ji, Dynam. Syst. **2005**, 225 (2005).

[19] http://docs.scipy.org/doc/numpy/reference/generated/numpy.linalg.slogdet.html

[20] M. Selig, M. R. Bell, H. Junklewitz, N. Oppermann, M. Reinecke, M. Greiner, C. Pachajoa, and T. A. Enßlin, Astron. Astrophys. **554**, A26 (2013).

[21] E. Jaynes and G. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003), chap. 12.4.

[22] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday *et al.* Planck Collaboration, Astron. Astrophys. **571**, A22 (2014).

[23] J. Martin, C. Ringeval, R. Trotta, and V. Vennin, J. Cosmol. Astropart. Phys. 03 (2014) 039.

[24] J. Skilling, AIP Conf. Proc. **735**, 395 (2004).

[25] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).

[26] T. Bayes, Phil. Trans. Roy. Soc. **53**, 370 (1763).

[27] S. L. Bridle, R. Crittenden, A. Melchiorri, M. P. Hobson, R. Kneissl, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **335**, 1193 (2002).

[28] T. A. Enßlin, H. Junklewitz, L. Winderling, M. Greiner, and M. Selig, Phys. Rev. E **90**, 043301 (2014).

[29] S. Dorn, T. A. Enßlin, M. Greiner, M. Selig, and V. Boehm, Phys. Rev. E **91**, 013311 (2015).

[30] C. M. Hirata and U. Seljak, Phys. Rev. D **67**, 043001 (2003).

[31] http://www.mpa-garching.mpg.de/ift/nifty

[32] T. A. Enßlin and M. Frommert, Phys. Rev. D **83**, 105014 (2011).